Supplementary Appendix

for: Revell, L. J. 2010. Phylogenetic signal and linear regression on species data. *Methods in Ecology & Evolution*.

In this article, I used a phylogenetic regression both with and without a flexible residual error structure in which off-diagonals of the error covariance matrix are multiplied by the parameter, λ, of Pagel (1999). These analyses can easily be performed using the `ape` and `nlme` package in R (Paradis *et al.* 2004; Pinheiro *et al.* 2009; R Development Core Team 2009). For instance, the standard phylogenetic regression, in which we assume a strict Brownian model for the residual error variances and covariances, can be fit as follows:

```
> require(ape)
> require(nlme)
> result<-gls(y~x,data=data,correlation=corBrownian(phy=tree),method="ML")
```

Here, `x` and `y` are column names in the data frame `data`. No column of 1.0s for the intercept is required. To construct a multivariable model, one simply needs to add additional columns to `data` and then fit: `y~x1+x2` etc. The results can be displayed using the command:

```
> summary(result)
```

Although this procedure will be satisfactory for most users, one can of course also fit the model using the equations of the main text, as follows:

```
> require(ape)
> C<-vcv.phylo(tree)
> n<-ncol(C)
> X<-as.matrix(cbind(matrix(1,n,1),data["x"]))
> colnames(X)<-c("ones","x")
> y<-as.matrix(data["y"])
> C<-C[rownames(y),rownames(y)]
> X<-X[rownames(y),]
> beta<-solve(t(X)%*%solve(C)%*%X)%*%(t(X)%*%solve(C)%*%y)
```

We can also compute our estimates of the variances and covariances of our parameter estimates, as follows:

```
> W<-as.double(t(y-X%*%beta)%*%solve(C)%*%(y-X%*%beta))/(n-2)*solve(t(X)%*%
solve(C)%*%X)
```

The standard error of the regression slope (`beta[2,1]`) is then given by:

```
> SE.beta1<-sqrt(W[2,2])
```

Phylogenetic regression with the correlation structure based on Pagel's λ (Pagel 1999), can also be performed easily in R. Here, we fit our model as follows:

```
> require(ape)
> require(nlme)
> result<-gls(y~x,data=data,correlation=corPagel(value=1,phy=tree,
```

```
fixed=FALSE),method="ML")
```

The command:

```
> summary(result)
```

now gives us not only the parameters of the regression model and their standard errors, but also the optimized value of λ.

As before, one can also fit this model using the equations of the main text, and the R optimization routine `optim()`. This is performed as follows:

```
> require(ape)
> C<-vcv.phylo(tree)
> C<-C[rownames(y),rownames(y)]
> likelihood.lambda<-function(lambda,y,X,C){
     V<-diag(diag(C))
     C<-C-V
     n<-nrow(C)
     C.lambda<-(V+lambda*C)
     beta<-solve(t(X)%*%solve(C.lambda)%*%X)%*%(t(X)%*%solve(C.lambda)%*%y)
     sig2e<-as.double((1/n)*(t(y-X%*%beta)%*%solve(C.lambda)%*%(y-X%*%beta)))
     logL<--(1/2)*t(y-X%*%beta)%*%solve(sig2e*C.lambda)%*%(y-X%*%beta)-(1/2)*
        determinant(sig2e*C.lambda,logarithm=TRUE)$modulus-(n/2)*log(2*pi)
     return(-logL)
}
> ML.opt<-optim(runif(1),likelihood.lambda,y=y,X=X,C=C,method="L-BFGS-B",
lower=0.0,upper=1.0)
> V<-diag(diag(C))
> C<-C-V
> n<-nrow(C)
> lambda<-ML.opt$par
> logL<--ML.opt$value
> C.lambda<-(V+lambda*C)
> beta<-solve(t(X)%*%solve(C.lambda)%*%X)%*%(t(X)%*%solve(C.lambda)%*%y)
> colnames(beta)<-c("beta")
> sig2e<-(1/n)*(t(y-X%*%beta)%*%solve(C.lambda)%*%(y-X%*%beta))
```

Here, `y` and `X` are as they were specified in preceding text. Fixing an upper and lower bound on the optimization is not strictly necessary, but may be helpful in obtaining sensible results. As before, we can compute the variance-covariance matrix of our parameter estimates, as follows:

```
> W<-as.double(t(y-X%*%beta)%*%solve(C.lambda)%*%(y-X%*%beta))/(n-2)*
solve(t(X)%*%solve(C.lambda)%*%X)
```

The standard error of the regression slope is then given by:

```
> SE.beta1<-sqrt(W[2,2])
```

In the present article I computed phylogenetic signal for individual traits using two methods. The first was by using Pagel's (1999) λ , but this time as a univariate measure of signal in a single trait. This can be performed using the same function, `likelihood.lambda()`, that was given above, as follows:

```
> C<-vcv.phylo(tree)
> C<-C[rownames(y),rownames(y)]
> n<-ncol(C)
> ones<-matrix(1,n,1)
> ML.opt<-optim(runif(1),likelihood.lambda,y=y,X=ones,C=C,method="L-BFGS-B",
lower=0.0,upper=1.0)
```

I also computed phylogenetic signal using the measure, *K*, proposed by Blomberg *et al.* (2003). *K* can be calculated as follows:

```
> C<-vcv.phylo(tree)
> C<-C[rownames(y),rownames(y)]
> n<-ncol(C)
> ones<-matrix(1,n,1)
> a<-as.double(solve(t(ones)%*%solve(C)%*% ones)%*%(t(ones)%*%solve(C)%*%y))
> K<-(t(y-a*ones)%*%(y-a*ones)/(t(y-a*ones) %*%solve(C)%*%(y-
a*ones)))/((sum(diag(C))-n/sum(solve(C)))/(n-1))
```

Appendix References

Pagel, M. (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289-290.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & the R Core team (2009) nlme: Linear and nonlinear mixed effects models. *R package version 3.1-96*.

R Development Core Team (2009) R: a language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna Austria.